AMENDMENTS TO THE CLAIMS:

This listing of the claims will replace all prior versions, and listings, of the claims in this application:

The Claims:

1. **(Currently Amended)** A method to process at least one text document, comprising:

partitioning text of each of the at least one text document into text tokens; ~~and~~

assigning semantic meaning to words of the partitioned text token, where assigning comprises applying a plurality of regular expressions, rules and a set of dictionaries,

where the set of dictionaries ~~comprising~~ is selected from the group consisting of:

a first collection of dictionaries consisting of a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments and

a second collection of dictionaries consisting of the common chemical prefix dictionary, the common chemical suffix dictionary to recognize chemical name fragments and a dictionary of stop words to eliminate erroneous chemical name fragments;

recognizing any substructures present in the chemical name fragments based at least in part on the semantic meaning assigned to words of the partitioned text token;

extracting keywords of the text document, where the keywords are associated with the recognized chemical name fragments and the substructures; ~~of the text document and~~

indexing the extracted keywords in a text index;

adding each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

determining structural connectivity information ~~of~~ within each of the recognized chemical name fragments and the substructures that do not contain a number;

indexing representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index, where indexing the representations comprises, for each of the at least one text document:

in a loop, testing each of the recognized chemical name fragments in ~~a first text document of~~ the at least one text document to see if the recognized chemical name fragment occurs in a dictionary of SMILES fragments, where if it does then a

SMILES expression for the recognized chemical name fragment ~~token~~ is added to the chemical substructure index,

then determining if the recognized chemical name fragment occurs in a MOL file dictionary, where if it does then a MOL file expression for the fragment token is added to the chemical substructure index, and ~~then~~

~~determining if there is a next text document of the at least one text document, where if there is a next text document then testing, as stated above, each of the recognized chemical name fragments in the next text document and where if there is no next text document the indexing is completed;~~

storing the text index in association with the chemical substructure index;

providing a graphical user interface to search the text index and the chemical substructure index, where the graphical user interface comprises a graphical list of substructures

where the search comprises first entering search terms comprising one or more chemical fragment names and then selecting graphical representations of one or more substructures, where the selecting comprises using the graphical user interface as a pointer to **[[a]]** the graphical list of substructures; and

receiving a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least one text document where there are found chemical compounds that contain the selected substructures, and connectivity specified by the one or more chemical fragment names and the selected substructures and where the search terms are found in the text index.

2. (**Currently Amended**) The method as in claim 1, wherein the search further comprises first entering search terms comprising the one or more chemical fragment names and entering at least one keyword, and where the search result is identifying at least one text document where there are found the at least one keyword, the chemical compounds that contain the selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

3. (Previously Presented) A method as in claim 1 performed by executing a computer program product.

4-6. (Canceled).

7. (Previously Presented) The method as in claim 1, where determining structural connectivity information comprises looking up recognized chemical name fragments and substructures in a structure dictionary.

8. (Canceled).

9-10. **(Canceled)**.

11. (Previously Presented) The method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

12. (Previously Presented) The method as in claim 1, where chemical name fragments are further recognized by using common chemical word endings.

13. (Previously Presented) The method as in claim 1, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed from between chemical name fragments as a function of context.

14. (Previously Presented) The method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

15. (Previously Presented) The method as in claim 14, where the punctuation comprises at least one of a parenthesis, a square bracket, a hyphen, a colon and a semi colon.

16. (Previously Presented) The method as in claim 14, where the characters comprise upper case C, O, R, N and H.

17. (Previously Presented) The method as in claim 14, where the characters comprise lower case xy, ene, ine, yl, ane and oic.

18. (**Currently Amended**) The method as in claim 1, comprising an initial step of tokenizing the text document to provide a sequence of tokens.

19. (**Currently Amended**) A system having at least one computer, comprising:

a tokenizer module and a token processing module comprised of computer instructions in data storage distributed across the at least one computer directing the at least one computer to partition text of each of at least one text document into text tokens and to assign semantic meaning to words of the partitioned text token by applying a plurality of regular expressions, rules and a set of dictionaries,

where the set of dictionaries ~~comprising~~ is selected from the group consisting of:

a first collection of dictionaries consisting of a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments and

a second collection of dictionaries consisting of the common chemical prefix dictionary, the common chemical suffix dictionary to recognize chemical name fragments and a dictionary of stop words to eliminate erroneous chemical name fragments;

the instructions of the token processing module directing the at least one computer to recognize any substructures present in the chemical name fragments based at least in part on the semantic meaning assigned to words of the partitioned text token;

the instructions of the token processing module directing the at least one computer to extract keywords of the text document, where the keywords are associated with the recognized chemical name fragments and the substructures ~~of the text document~~ and to index the extracted keywords in a text index;

the instructions of the token processing module directing the at least one computer to add each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

the instructions of the token processing module directing the at least one computer to , for each of the at least one text document, determine structural connectivity information ~~of~~ within each of the recognized chemical name fragments and the substructures that do not contain a number, and to index representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index,

where indexing the representations comprises:

in a loop, testing each of the recognized chemical name fragments in ~~a first text document of~~ the at least one text document to see if the recognized chemical name fragment occurs in a dictionary of SMILES fragments, where if it does then a SMILES expression for the recognized chemical name fragment ~~token~~ is added to the chemical substructure index,

then determining if the recognized chemical name fragment occurs in a MOL file dictionary, where if it does then a MOL file expression for the fragment token is added to the chemical substructure index, and ~~then~~

~~determining if there is a next text document of the at least one text document, where if there is a next text document then testing, as stated above, each of the recognized chemical name fragments in the next text document and where if there is no next text document the indexing is completed;~~

the instructions of the token processing module directing the at least one computer to store the text index in association with the chemical substructure index;

a searcher module comprised of computer instructions distributed across the at least one computer and a graphical user interface comprised of a display configured to display a graphical list of substructures and a keyboard connected to a computer of the at least one computer directing the at least one computer to search the text index and the chemical substructure index, where the search comprises first entering one or more chemical fragment names and then selecting graphical representations of one or more substructures, where the selecting comprises using the graphical user interface as a pointer to **[[a]]** the graphical list of substructures; and

the graphical user interface configured to receive a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least

one text document where there are found chemical compounds that contain the selected substructures, and connectivity specified by the one or more chemical fragment names and the selected substructures and where the search terms are found in the text index.

20. (**Currently Amended**) The system as in claim 19, wherein the search further comprises first entering the one or more chemical fragment names and additionally entering at least one keyword, and where the search result is identifying at least one text document where there are found the at least one keyword, the chemical compounds that contain the selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

21-24. (Canceled).

25. (Previously Presented) The system as in claim 19, where the instructions of said token processing module that directs the at least one computer to determine the structural connectivity information further directs the at least one computer to look up recognized fragments and substructures in a structure dictionary.

26. (Canceled).

27-28. (**Canceled**).

29. (Previously Presented) The system as in claim 19, further comprising the instructions of said token processing module further directs the at least one computer to filter recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

30. (Previously Presented) The system as in claim 19, where the instructions of the tokenizer module further directs the at least one computer to recognize chemical name fragments by using common chemical word endings.

31. (Previously Presented) The system as in claim 19, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed from between chemical name fragments as a function of context.

32. (Previously Presented) The system as in claim 19, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

33. (Previously Presented) The system as in claim 32, where the punctuation comprises at least one of a parenthesis, a square bracket, a hyphen, a colon and a semi colon.

34. (Previously Presented) The system as in claim 32, where the characters comprise upper case C, O, R, N and H.

35. (Previously Presented) The system as in claim 32, where the characters comprise lower case xy, ene, ine, yl, ane and oic.

36. (**Currently Amended**) The system as in claim 19, further comprising an input tokenizer module comprised of computer instructions directing the at least one computer to receive text documents to be processed to provide a sequence of tokens.

37-41. (Canceled).

42. (**Canceled**).

43. (**Currently Amended**) A system comprising a plurality of computers at least two of which are coupled together through a data communications network, said system comprising:

a tokenizer and a token processing module comprised of computer instructions in data storage distributed across the plurality of computers directing the plurality of computers to parse text of each of at least one text document into text tokens and assign semantic meaning to words of the parsed ~~sentences~~ text tokens,

where assigning comprises applying a plurality of regular expressions, rules and a set of dictionaries,

where the set of dictionaries ~~comprising~~ is selected from the group consisting of:

a first collection of dictionaries consisting of a common chemical prefix dictionary and a common chemical suffix dictionary to recognize chemical name fragments[[;]] and

a second collection of dictionaries consisting of the common chemical prefix dictionary, the common chemical suffix dictionary to recognize chemical name fragments and a dictionary of stop words to eliminate erroneous chemical name fragments;

the instructions of the token processing module directing the plurality of computers to recognize any substructures present in the chemical name fragments based at least in part on the semantic meaning assigned to words of the partitioned text token;

the instructions of the token processing module directing the plurality of computers to extract keywords of the text document, where the keywords are associated with the recognized chemical name fragments and the substructures ~~of the text document~~ and to index the extracted keywords in a text index;

the instructions of the token processing module directing the plurality of computers to add each of the recognized chemical name fragments and the substructures that do not contain a number to the text index;

the instructions of the token processing module directing the plurality of computers to, for each of the at least one text document, determine structural connectivity information ~~of~~ within each of the recognized chemical name fragments and the substructures that do not contain a number;

the instructions of the token processing module directing the plurality of computers to index representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables of a chemical substructure index,

where indexing the representations comprises:

in a loop, testing each of the recognized chemical name fragments in ~~a first text document of~~ the at least one text document to see if the recognized chemical name fragment occurs in a dictionary of SMILES fragments, where if it does then a

SMILES expression for the recognized chemical name fragment ~~token~~ is added to the chemical substructure index,

then determining if the recognized chemical name fragment occurs in a MOL file dictionary, where if it does then a MOL file expression for the fragment token is added to the chemical substructure index, and ~~then~~

~~determining if there is a next text document of the at least one text document, where if there is a next text document then testing, as stated above, each of the recognized chemical name fragments in the next text document and where if there is no next text document the indexing is completed;~~

~~the instructions of the token processing module directing the plurality of computers to store~~ storing the text index in association with the chemical substructure index;

a searcher module comprised of computer instructions distributed across the plurality of computers and a graphical user interface comprised of a display configured to display a graphical list of substructures and a keyboard connected to a computer of the plurality of computers directing the plurality of computers to search the text index and the chemical substructure index, where the search comprises first entering search terms comprising one or more chemical fragment names and then selecting graphical representations of one or more substructures, where the selecting comprises using the graphical user interface as a pointer to [[a]] the graphical list of substructures; and

the graphical user interface configured to receive a search result, where the search result is an intersection of the chemical substructure index and the text index, identifying at least one text document where there are found chemical compounds that contain a reference to the search terms and the one or more substructures and where the search terms are found in the text index.

44. **(Currently Amended)** The system as in claim 43, wherein the search further comprises first entering the one or more chemical fragment names and additionally entering at least one keyword, and where the search result is identifying at least one text document where there are found the at least one keyword, the chemical compounds that contain the

selected substructures, and the connectivity specified by the one or more chemical fragment names and the selected substructures.

45. (Previously Presented) The system as in claim 43, where the instructions of said token processing module further direct the plurality of computers to look up recognized fragments and substructures in a structure dictionary.

46. (Canceled).